The cognitive decision space of trust: an exploratory study of image trustworthiness and the propensity to deceive

L. Fell, P.D. Bruza, S.K. Devitt
Information Systems School
Queensland University of Technology, Brisbane, Australia
p.bruza@qut.edu.au

G. Oliver, M. Gradwell
Victoria University of Wellington, New Zealand

H. Partridge
University of Southern Queensland, Toowoomba, Australia

In an age where any digital image can be manipulated, studying how and why people trust images as well as how likely people are to endorse deceptive images has become a topic of increasing importance. But, how is it human beings decide whether they trust and image, or not? This article attempts to **shed light on the cognitive decision space of trust by means of two experiments. The goal of the first experiment was to induce the dimensions underpinning decisions of trust in relation to images. The goal of the second experiment was to investigate the propensity for subjects to deceive with images in conditions of both high and low levels of deception**. **The first experiment revealed four dimensions that determined the level of trust in an image: the features of the image, its content, its source, and the participants' own background knowledge. The second experiment suggests that there is propensity for human subjects to deceive with images.**

Introduction

In an age where any digital image can be manipulated, studying how and why people trust images as well as how likely people are to endorse deceptive images has become a topic of increasing importance. Mapping the dimensions of cognitive decision space may shed light on how human beings decide whether they trust an image, or not and the propensity for human beings to deceive with images. A cognitive decision space framework provides tools to analyse, categorise and predict trust likelihoods for images.

The framework draws on the methodological approach taken to map out the decision space regarding document relevance. Over three decades, qualitative studies have identified cognitive dimensions of relevance, which have exhibited an encouraging degree of inter-subjective agreement (Schamber et al., 1990; Barry, 1994; Mizzaro, 1997; Borlund, 2003). For example, a recent study examined how users determined which list of search engine results (in the form of document captions) they preferred over another using five dimensions of relevance: "topicality," (how well the caption was topically related to the user query), "freshness" (currency), "authority" (credibility), "caption quality," and "diversity" (Kim et al., 2013). In this article, we focus on "trust" rather than "relevance", using definitions of the concept of trust from scholars in cognitive science, archival and information science.

Trust, Images and Manipulation

Trust is a philosophical ideal that has undergone many paradigm shifts over the past few centuries (MacNeil, 2001, p. 37). Trust implies a standard of moral certainty regarding how facts can be established as a result of empirical enquiry or reflective equilibrium. Judgement of trust is pivotal in the world of records where value is established because of the evidential aspect of records. Ensuring a high evidential value of records is the domain of diplomatics, which conceives of evidence as inference (MacNeil, 2001, p. 39). Diplomatics is a methodology by which documents (including text, images and multi-media) are evaluated to ensure that records are as trustworthy as possible. It is these ideas of records as evidence, and evidence as inference, that has largely shaped the direction of recordkeeping during the nineteenth and twentieth centuries.

Outcomes from the Canadian-led InterPARES (International Research into the Preservation of Authentic Records in Electronic Systems) research projects concluded that trustworthiness of a record can be defined by three key aspects. These are "reliability", "accuracy" and "authenticity" (Duranti & Rogers, 2012, p. 525; InterPARES, 2008; MacNeil, 2001, p. 40). An expansion of these terms identifies that:

- Reliability is the trustworthiness of a record as a statement of fact and it is capable of standing for the action to which it attests. The reliability of a record can be ascertained based on the competence of its author, its completeness, and the controls on its creation.
- Accuracy is the degree to which data, information, documents or records are precise, correct, truthful, free of error or distortion, or pertinent to the matter. This is based on the former and on the controls on the recording of content and transmission.
- Authenticity is defined as the trustworthiness of a record as a record. This means that the record is what it purports to be, free from tampering or corruption. What the record professes in origin or authorship is genuine.

Kelton et al.'s (2008, p370) model of trust in information comprises similar features:

- Accuracy: the extent to which information is free from error
- Objectivity: the balance of content
- Validity: the use of responsible and accepted practices such as the soundness of the methods used, the inclusion of verifiable data, and the appropriate citation of sources
- Stability: the persistence of information, both its presence and contents

Donaldson and Conway (2015) conducted a qualitative study of user conceptions of trust of archival documents using Kelton et al.,'s model as a point of departure. One of the premises of this study is shared by the present authors, namely, archival studies has "tended to treat the end user, when invoked at all, as the recipient of 'propertied' information, rather than as participants in the formation of trustworthiness".

Trust in digital contexts concerns the three values of reliability, accuracy and authenticity and has been researched under the banner of 'eTrust' (Taddeo & Floridi, 2011; Taddeo, 2009). In this research we ask: What are the dimensions underpinning decisions of eTrust in relation to images? The first point of note is that because digital contexts are less curated than traditional domains of record management, trust is defined against risk. That is, trust is accorded against the risk of not trusting, and the risk of trusting an untrustworthy or misleading source.

Risk is intimately tied to honesty. The risk of a dishonest image is that it misleads the viewer with regards to what it represents, potentially causing harm via false beliefs. Professional photographers have been particularly active with regards to issues of eTrust and have proposed a universal ethical protocol. The hope is that an ethical protocol may contribute to the trustworthiness of images that are constantly seen in the media. For example, The National Press Photographers Association (NPPA) in the United States has a primary goal of, "the faithful and comprehensive depiction of the subject at hand (National Press Photography Association, 2012)." They see their work as having historic value and their values indicate that there should be no manipulation before or after the photo has been taken that could mislead viewers or misrepresent subjects. Nevertheless, though standards exist, there is no strict enforcement of their values. Often news agencies create their own sets of ethics for their photographers to abide by. The Los Angeles Times is one such example: their guidelines are more pragmatic than that of the NPPA and go on to detail exactly what adjustments are and are not acceptable. Any artistic renderings of images are to be clearly labelled "photo illustration" (Los Angeles Times, 2005). If there is to be a universal protocol, then it has been proposed that there must be a definition of photographs under categories in the same manner in which texts are defined. This would place photographs into genres such as fiction and non-fiction, or editorializing and reportage (Roberts & Webber, 1999, p. 3). Within this it would be possible to use specific terminology to differentiate manipulated photographs from the untouched. While this could be an effective tool for live records, it does not help so much once those images are subsumed into the archives.

It is from concerns regarding the edge of what is acceptable or not acceptable that reputed organizations with image archives such Getty Images (www.gettimages.com) and Reuters (http://www.reuters.com/) have a zero tolerance policy on photo manipulations (Lum, 2010, July 5; Lum 2010, July 19). A photographer who modified a golfing image to remove a background bystander was terminated by Getty Images in accordance with this policy (Lum, 2010, July 19). Similarly, a Pulitzer prize-winning photographer was fired after he had admitted altering an image of the conflict in Syria by photoshopping a camera out of the image[1]. In both cases the stance taken is that the image be a totally true and accurate depiction of reality regardless of how innocuous the alteration. Therefore, we define a decision on the trustworthiness of an image to *be a decision on whether the image is an accurate representation of a situation, person or object.* Naturally, much hinges on how accuracy is interpreted and where the subject sets the threshold for the image being "accurate enough". For example, a subject might still judge the Pulitzer prize winner's photograph as being accurate if they knew the camera had been

[1] http://www.dailymail.co.uk/news/article-2544662/Pulitzer-Prize-winning-photographer-fired-admitting-doctored-Syrian-war-rebel-picture-photoshopping-camera-original-image.html

photoshopped out, simply because the object erased did not impinge on its resemblance to an actual war scene in Syria. Matt Carlson (2009) wrote that, "While photography's fidelity to the real world has long been subject to speculation within both the academy and journalism, the diffusion of digital imaging technologies and software raises further questions concerning manipulation and alteration (p. 126)." The reality is that manipulated digital images are now an everyday occurrence, but to what extent are they deceitful? Deceit implies intent on the behalf of the photographer or the manipulator of the image, but there are varying degrees of manipulation, and different viewer reactions. In short, judgments of trust involve complex contextual factors that affect the perception and processing of images.

Greenberg (2013) highlights two cognitive processes that are in operation when a person views an image: those that evaluate the content, resemblance and reference of an image and processes that evaluate the geometric and artistic depiction of an image with regards to reality. Because these processes operate implicitly and pre-consciously, it is possible that they confound in some way if an image is ambiguous. That is, if viewers are challenged with regards to the content, resemblance or reference of an image, it might precipitate a challenge to the geometric or artistic representation of the image and vice versa. It turns out that visual fluency is an important factor. The concept of visual fluency is based on the principle that any visual stimulus requires cognitive work to process, the more work required, the less fluent the process. Cognitive work includes the evaluation of: content, resemblance, reference of an image; geometric and artistic depictions. Images that cohere with background beliefs on any of these factors are more easily processed than properties that surprise or confuse us. The amount of cognitive work is reflected in the speed and accuracy of visual processing as well as in the subjective experience of ease or difficulty of visual judgments (Jacoby, Kelley, & Dywan, 1989; Winkielman, Schwarz, Reber, & Fazendeiro, 2003). Factors such as blurriness (Shah & Oppenheimer, 2007) and contrast (Reber & Schwarz, 1999) can affect visual fluency. If the visual fluency hypothesis is right, then manipulated photos are less detectable the more they conform to largely unconscious rules of visual fluency. Interestingly, people are typically unaware why a given stimulus is easy to process, so their judgments can be manipulated. Ease of visual processing results in an illusion of truth, perhaps because perceptual fluency elicits a feeling of familiarity (Winkielman et al., 2003, p. 7)—and hence trust. As Shah & Oppenheimer (2007) note, "…with so much information available, how do we decide which cues to weight most heavily when we make visual veracity decisions?" (p. 371).

When images include human faces, a slew of additional mechanisms come into play that might confound overall trust judgments of images. Rather than a judgment of accuracy, judging trust from facial appearance triggers basic approach/avoidance responses in social situations (Todorov, 2008). Facial trust judgments, like judgements of attractiveness, are an efficient heuristic to social decision-making (Willis & Todorov, 2006), that is, it only takes 33ms to discriminate between a trustworthy and untrustworthy-looking face (Todorov, 2008). Trustworthy faces correlate with features identified with happiness; where as untrustworthy faces correlate with features identified with anger (Todorov & Duchaine, 2008). Additionally, typical faces, evoking familiarity, are more trusted than atypical faces (Sofer, Dotsch, Wigboldus, Todorov, 2014). The effect of typicality on trust is found not only for faces, but also in music (Repp, 1997), colours (Martindale & Moore,

1988) and nonface objects (Halberstadt & Rhodes, 2003). Typicality research comports with the visual fluency research mentioned above where the more familiar a stimulus, the easier it is to process, the more positive affect created and the more trusted it is (Winkielman, Halberstadt, Fazendeiro & Catty, 2006). It does not matter what sort of task one undertakes, processing fluency generates trust (Alter & Oppenheimer, 2009).

Given the complexity of background cognitive processes which may confound decisions of trust, it is not surprising that in the few studies that have been conducted, human beings are not adept at making robust decisions about the trustworthiness of images. One recent study rated human subjects at being "poor to moderate" in their ability to detect manipulated images, and "poor" at identifying what part of the image had been manipulated (Caldwell et al., 2015). Given this difficulty in detecting manipulation, it is possible that the subject's personality type might sway the decision one way or the other. For example, those with personality traits of openness, agreeableness or emotional stability may exhibit a propensity to trust images whilst those who are conscientious may be predisposed to be critical and have a propensity to distrust. It seems reasonable to assume that when the manipulation does not sufficiently disturb resemblance, that other factors may be employed in the decision making process. As noted above, "authority" (credibility) has been one dimension shown to affect decision space around document relevance. The source of information has also been shown to affect persuasion (Smith, Houwer & Nosek, 2013, Pornpitakpan, 2004). If the credibility embedded in a particular source has the ability to change people's implicit evaluations, this could also be the case with trust. For example, if the source of an image is a reputed institution, the subject might be more willing to trust the image than if it was encountered on social media. Similarly, a journalist may elicit more trust than a blogger. Indeed, journalists are primarily perceived to have higher standards of credibility given the strict professional standards imposed on them, which is lacking in the blogging community (Davis, 2008). Nick Denton, owner of a series of web blogs told the New York Times, "I think it's implicit in the way that a Web site is produced that our standards of accuracy are lower" (Bosman, 2004, Section 9, p. 10).

Importantly, trust is not merely a cognitive phenomenon, but a sociological one too. The sociological aspect of trusting internet images is particularly acute because frequently images are not associated with reputational source such as Reuters or Getty Images, but appear de-contexualised to agents through anonymising sources such as image search engine results (e.g. Google image search), photo sites (e.g. Flickr), clickbait aggregator sites (e.g. Dose), or personal websites (e.g. Wordpress) without attribution. The anonymity of internet images removes relational aspects of trust that define ordinary social interactions (Cook & Gerbasi 2009). Removing intentionality from the images also reduces them to a Popperian 'third world', or 'objective' documents (Swanson,1986; Popper, 1972). Such images rely on an agent's best-guess regarding accuracy and intention of the photographer/artist, and thus reduces the epistemic 'benefit of the doubt' that thrives with reasonable latitudes of interpretation between cooperative agents. Indeed, the very possibility of etrust under such conditions has been questioned (Nissenbaum, 2001).

In addition to decisions of trustworthiness, there is the question to what degree subjects are willing to deceive with images. The classic example of this is where

blemishes are removed from personal photographs in order to make them look better. In addition, media outlets are regularly exposed in using images or videos to depict a story, but where these pertain to a like, but different event. This raises the question about what factors influence the decision threshold above which subjects are willing to deceive.

With preceding as background, this article will present two experiments aimed at clarifying salient aspects of the cognitive space which frames decisions of trust in relation to images, as well as exploring the propensity of subjects to deceive with images.

Experiment 1: Cognitive dimensions of trust

The goal of the first experiment is to induce the dimensions that underpin decisions of trust in relation to images. A qualitative methodology was employed similar to studies which clarified the dimensions underpinning decisions of relevance. In addition, it is hypothesised that a reputed source accompanied with an image will produce higher trustworthiness ratings than a social media source or a lack of any source. We also predict that people high on scales of the personality traits of "openness", "agreeableness" and "emotional stability" would be more likely to provide higher overall trustworthiness ratings, whilst those high on "conscientiousness" would provide lower overall ratings.

Participants

Participants consisted of 87 workers using the online crowdsourcing platform, *Amazon Mechanical Turk* (AMT). This is a platform that enables researchers, amongst others, to post experiments and surveys in a form called a HIT (Human Intelligence Task) on a website available to thousands of potential participants to view and complete. Participants involved in the present study were able to view any HIT before agreeing to participate, and were paid a small amount (such as 50 – 60c) per HIT. AMT allows workers of a certain skill, ability or reputation to be specified for a HIT. In this experiment, workers of at least 95% or greater approval rating were selected to balance worker quality with the need to attract a sufficient number of participants. No demographic data was taken of participants, however, the general demographics of AMT workers are generally known. Around 50% are from the United States, 40% from India, and 10% from other countries (Ipeirotis & Panagiotis, 2010). Workers using this platform are predominantly female if residing in the United States, and predominantly male if residing in India (Ipeirotis & Panagiotis, 2010). Because the experiments were presented in English, it was assumed that workers choosing to participate would be proficient in this language. Data collected from participants who did not respond to written segments of the experiment in English, or who were deemed not to have understood instructions based on the relevancy/quality of their responses, were excluded from the study. In total, 3 workers were excluded.

Materials

Ten Item Personality Inventory (TIPI) developed by Gosling et al (2003) was designed as a short measure of the five-factor model of personality. The five-factor model stipulates that personality can be broken into five scales; openness to experiences, conscientiousness, extraversion, agreeableness and emotional stability.

Participants were asked to what extent each item reflects how they see themselves on a 7-point scale. Each personality trait had both a positive and reverse-scored item, for example, the two items measuring extraversion were *extraverted, enthusiastic*, and *reserved, quiet* (reverse-scored). Gosling, Rentfrow & Swann (2003) found the measure to have stong test-retest reliability (r = .72) as well as convergent validity (mean r = .77).

Images were obtained using a *Google images* search, and were freely available for use. All images used were chosen because they represented an unusual or unexpected depiction of the subject they portrayed. The group included a mix of digitally altered and unaltered images. The aim of image selection was to present images that were difficult to process due to their visual disfluency, therefore triggering a predisposition to distrust, whilst containing a mixture of geometrically 'real' and 'fake' images to assess how these influencers of trust interact. The following images were used, in order, a photograph of Russian President Vladmir Putin (unaltered), a photograph of a frill shark (unaltered), a picture of a man running from an explosion (digitally altered), a photograph of a mountain with an image taken of deep space by the Hubble Telescope superimposed in the background (digitally altered), and a photograph of a train derailment at Montparnasse Station in 1895 (unaltered). Vladmir Putin, for example, is not often seen with a smile, and a frill shark is far from prototypical of its species.

1.

2.

3.

4.

5.

Procedure

Participants were instructed to peruse each HIT before agreeing to participate. Each HIT began with the 10-item personality questionnaire, followed by a definition of trustworthiness and a set of questions. The definition of trustworthiness given to participants was as follows:

Trustworthiness can be defined as an accurate representation of a situation, person or object

This was followed by 5 questions, each of which began with an image. Each image was contained within a red border and presented in the same order in each HIT. There were 3 experimental conditions; two contained sources underneath each image (outside the image's red border), and one lacked any source. Of the two conditions assigning sources, one contained a source of social media origin (Facebook), whilst the other was a reputed source (either the Museum of Natural History in South Africa, or the Museum of Modern History in South Africa, depending on the nature of the image). Following each image was a question, which read:

Taking into account the image itself*, please indicate the level to which you judge the* trustworthiness *of the above image based on features* inside the red box.

This was followed by a Likert scale response system ranging from *very untrustworthy* (1) to *very trustworthy* (5) as well as a textbox asking participants to explain the reasons behind their decisions. After completion of the experiment, participants were given the opportunity to choose to submit their HITs.

Analysis

After all data had been gathered, qualitative data was coded using an axial coding method, with focus placed on themes of the features bearing on decisions of trust. Coding revealed four themes:

1. Features of the image itself (e.g. "*The person in the foreground does not seem to blend with the image in the background"*)
2. Content/Subject of the image (e.g. "*All wild animals are untrustworthy. Also look at his teeth. He acts on instinct alone."*)
3. Source below the image (e.g. *"it is from a museum, so it seems to be trustworthy"*)
4. Prior knowledge (e.g. "*Could be faked, but the auroras are spectacular and always look fake."*)

The results of this coding were used to create the categorical variable, *Decision Basis*. The effect of this, as well as the condition variable on the dependant variable of trustworthiness score, were analysed using the GLM procedure in SPSS. The

influences of personality factors, as well as the effects of the experimental conditions on trustworthiness ratings were assessed using ANOVAs.

Results

A one way ANOVA revealed significant effects of type of source on ratings of trustworthiness for the following images: Putin, $F(2,84) = 4.25$, $p = .017$, frill shark, $F(2,84) = 6.79$, $p = .002$, and mountain and sky, $F(2,84) = 3.13$, $p = .049$. Contrast tests were conducted on significant main effects. Contrasts revealed a significant difference between the Facebook and Museum sources for the Putin, $t(84) = 2.61$, $p = .01$, frill shark, $t(84) = 3.37$, $p = <.01$, and train images, $t(83) = 2.09$, $p = .04$, with the more reputed Museum source yielding higher trustworthiness ratings than the less reputed Facebook source. However, this difference was not found for the mountain and sky image, $t(84) = .93$, $p = .36$. For the mountain and sky image the differences between no source and both sources were significant $t(84) = 2.32$, $p = .02$, as well as between the Facebook source and no source $t(84) = 2.48$, $p = .02$, but not the difference between Museum source and no source $t(84) = 1.55$, $p = .13$. Upon examining the means, this appears to be due to the fact that the Facebook source yields a slightly higher mean rating of trustworthiness than the Museum source, with no source yielding much lower ratings than either.

One-way ANOVAs were conducted to find the effects of the basis of decision on decisions made on trustworthiness for each image.

	Features	Subject	Source	Prior Knowledge
Putin	14	83.7	1.2	1.2
Frill Shark	28.2	60	2.4	9.4
Explosion	63.5	34.1	1.2	1.2
Mountain	51.8	37.6	1.2	9.4
Train	38.8	27.1	1.2	32.9

Table 1 – Percentages of decision basis used by image

There was a significant relationship between the answers that participants gave and the reasons they reported for those answers for Putin $F(3, 82) = 11.86$, $p < .0001$, frill shark $F(3, 81) = 4.28$, $p = .007$, explosion $F(3, 81) = 3.09$, $p = .032$, mountain and sky $F(3, 81) = 11.82$, $p < .0001$, and train $F(3, 81) = 7.14$, $p < .0001$.

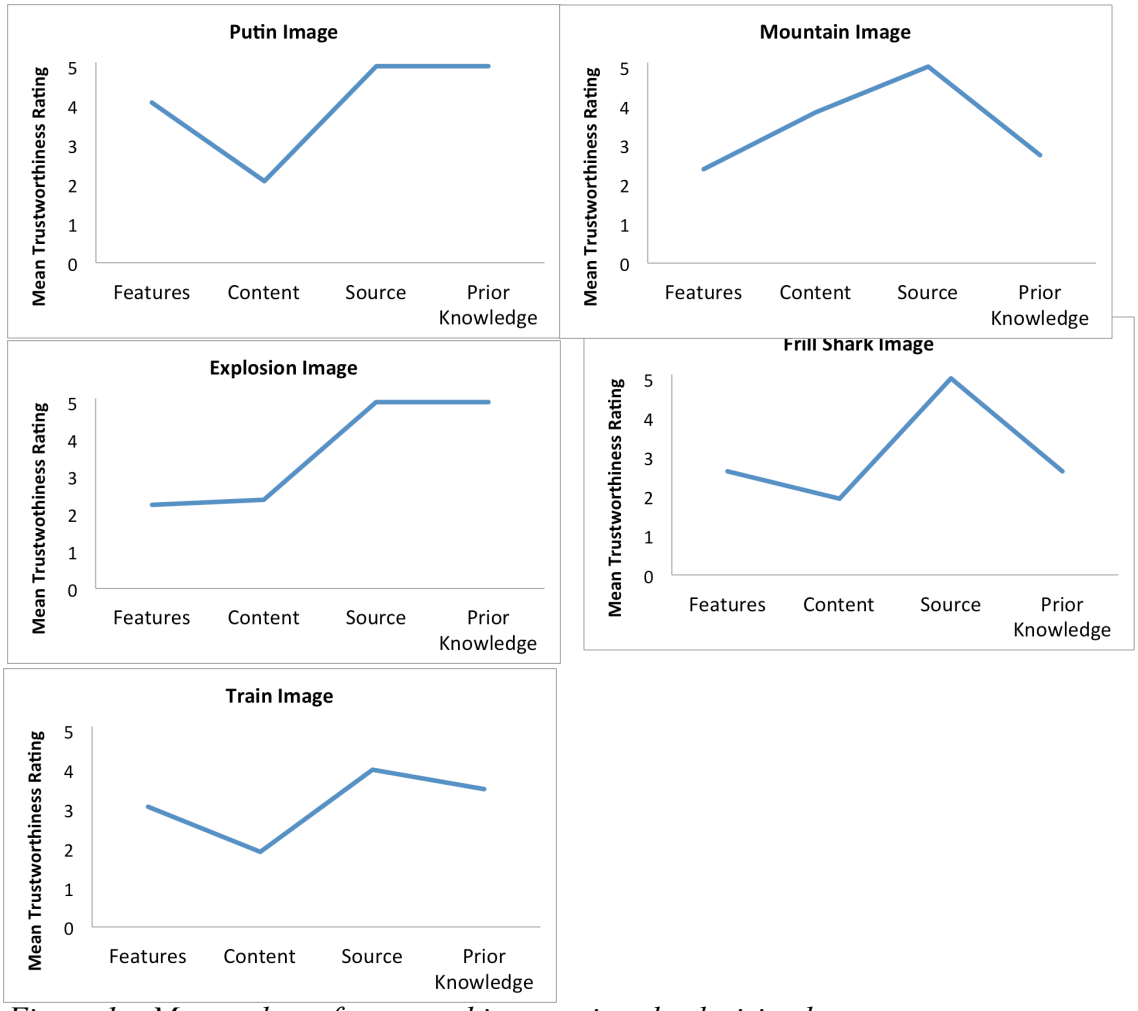


Figure 1 – Means plots of trustworthiness ratings by decision bases.

Chi Square analyses were conducted to determine whether certain answers were more affected by the subject of an image than others. To conduct this test, trustworthiness ratings were separated into 'trust' and 'not trust' answers by coding Likert scale responses from 1 to 2 to 'not trust' and 4 to 5 to 'trust'. For the Putin image, 98% of the participants who did not trust the image had based their answer on the subject of the image, whilst this was only 52% in people who trusted the image. Participants were significantly more likely to refer to the subject of this image as the reasons for their trust judgement if they distrusted the image itself, $c^2(2, 86) = 26.90$, $p = <.001$.

All ratings of trustworthiness were averaged to create a total score of propensity to trust. This was significantly positively correlated with the personality variable of Emotional Stability $t(1, 85) = 4.158$, $p=.045$.

Discussion

There was a significant effect of source on participants' ratings of trustworthiness for all but the explosion image condition. They all reflected that people generally trusted the museum sources over both Facebook and a lack of source. The train image showed only a significance between Facebook and museum sources, with no source

falling somewhere in between, showing that people trust Facebook even less than no source at all. These results were unsurprising, as we expected rational thought to conclude that a museum would present more credible source of images than social media.

However, this was not the case in every image. In the mountain and sky image, participants were significantly more likely to trust the image given a Facebook source compared to no source at all. The same was not true for the museum source. This unlikely result could be due to the actual source of the image used. It was a photoshopped image that had been circulated in social media. It is likely that the familiarity of the image in that context may have elicited more trust than seeing it with a museum source. This provides an interesting avenue for future research in finding the degree to which context familiarity affects people's trust in a sourced image.

Despite participants' apparently high reliance on source as a means to judge the images' trustworthiness, their self-reports did not match this at all. Averaging across images, only 1.38% of participants mentioned the source of an image as a reason for making their judgement. This, relatively concrete, means of judging the accuracy and originality of an image appears to be left largely unarticulated, and gives way to the more abstract means of determining whether an image portrays an expected representation in the outward expression of a trustworthiness judgement. The majority of participants in this case appear to be making *implicit* judgements based on the source and alteration indicators for making trust/don't trust judgements, however, are largely explaining these judgements by calling on more explicit challenges to their resemblance expectations of the subject of each image.

Contrasts revealed that, in all images, decisions based on the source of an image were linked to significantly higher trust ratings than other bases of decisions. In addition, participants basing their decisions on the subject matter of an image rated almost all images as significantly lower than other decision bases. The exception was the mountain image. The majority of qualitative data for participants choosing to base their decision on the subject matter revealed that participants were describing the aesthetics of the image. Speaking about prior knowledge was associated with significantly higher ratings of trustworthiness for the Putin, explosion and train images, but was linked with significantly lower trustworthiness ratings for the frill shark and mountain images.

In the Putin and Frill Shark images, the most common reasoning given was the subject of the image, where most were distrusting the personality of Putin and the menacing-look of the shark. This relates to research on visual fluency whereby an untrustworthy subject (whether due to a known political past, or large set of teeth), eliciting a negative response (Todorov, 2008), may impede visual fluency, thus leading to a judgement of distrust of the medium through which this untrustworthy subject was brought – the image itself. This, coupled with the non-representative depiction of each appears to have led participants to substantially distrust both images simply because they distrust the subject portrayed in the image. This is a curious finding.

In the explosion image, comments regarding the features of the image take precedence. Participants were noticing the signs within the image that it had been manipulated, so it is unsurprising that this reasoning was most common when determining trustworthiness. The same was true of the mountain image, in addition to the high percentage of participants discussing the aesthetics of this image to judge trustworthiness. Over a third of participants used prior knowledge to determine trustworthiness of the train image, which appears to show participants who were familiar with the incident. Most of the remaining participants were split between speaking about the unlikeliness of the event portrayed in the image and features of the image indicating that it may be altered. Given the high incidence of participants using the features of an unaltered image to determine its trustworthiness, it may suggest that people search for errors in image manipulation to confirm what they may not believe the image portrayed.

Speaking about prior knowledge was associated with significantly higher ratings of trustworthiness for the Putin, explosion and train images, but was linked with significantly lower trustworthiness ratings for the frill shark and mountain images. This appears to be due to participants often basing decisions on their unfamiliarity with a creature similar in appearance to the odd-looking frill shark. In the mountain image, participants rated trustworthiness as lower when they had seen the image before knowing that it had been photoshopped. The fact that basing decisions on source appeared to co-occur with higher ratings of trust could be due to a credible source overriding other possible untrustworthy contents of an image and thus determining the decision. This supports our quantitative findings that source significantly effects the trustworthiness ratings of images. What is interesting is the comparatively low incidence of participants discussing source as the basis for their decisions. This suggests that the source of an image does not need to be forefront in the mind to affect decisions of image trustworthiness.

In terms of personality, it was found that emotional stability was positively correlated with propensity to trust the images. The only other personality measure that yielded a significant result was in the explosion image, where the more extraverted a person was, the more likely they were to rate this image as trustworthy. This could be because the stimulating subject matter appealed to extraverts more so than the rest of the images. This is supported by the finding that the degree to which someone likes something positively impacts a person's trust in it (Doney & Cannon, 1997).

Experiment 2: The propensity to deceive with images

The goal of this experiment was to investigate the propensity for subjects to deceive with images. To this end, scenarios placed personal ethics in a tension with the severity of the deception. Ethical positions were established via roles: Journalist (assumed high ethics) vs. Blogger (assumed low ethics) and degree of purported deception with an image (high degree vs. low degree).

We hypothesized that participants assigned to the role of Journalist would have a lower propensity to deceive than people assigned to be a Blogger in their decisions to use/not use each image. Additionally, we hypothesized that the propensity to deceive would be higher in the low deception condition regardless of the role.

Participants

Participants consisted of 122 workers using the online crowdsourcing platform, *Amazon Mechanical Turk* (AMT). Participants involved in the present study were able to view any HIT before agreeing to participate, and were paid a small amount (such as 50 – 60c) to per HIT. As in Experiment 1, workers with at least 95% approval rating were recruited. No demographic data was taken from the participants. Data collected from participants who did not respond to written segments of the experiment in English, or who were deemed not to have understood instructions based on the relevancy/quality of their responses, were excluded from the study. In total, 2 workers were excluded.

Materials

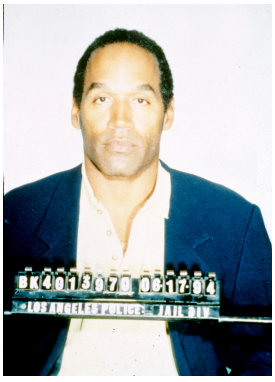
Images were obtained using a *Google images* search, and were freely available for use. The following images were used, in order, a photograph of the aftermath of a natural disaster in the Philippines, soldiers running toward a helicopter, tourists in front of Machu Picchu, and a mug shot of OJ Simpson, either original, or photoshopped. These images were chosen as they could be generalisable to each set of scenarios.

1.

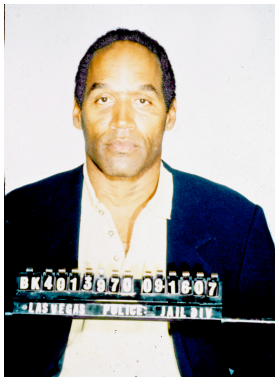
2.

3.

4a.



4b.



Procedure

Each HIT began with the following instructions:

For the following questions, you will be given a role and asked to explain what you would do in each situation. Please keep in mind this: It takes the photographer providing you with your photos 24 hours to produce each photo they offer to you. When you are making a living out of writing, you will be most successful if you are the first to publish a story, and you must be careful that rivals do not beat you to it. You must make each of the following decisions based on what you think you would do, along with the consideration of the time constraints present in your role-play occupation.

This was followed by 4 scenarios, each of which contained a description and an accompanying image. The experiment was designed as a between subjects 2 X 2 design. Two variables were created by manipulating each scenario. One variable was role, with half of the conditions requiring participants to imagine they were journalists, whilst participants in the other half were allocated the role of bloggers. The second variable was the severity of the deception of each image; high and low. Severity was manipulated by altering the description of the scenario. For example, one scenario asked that the participant imagine that they were a journalist (or blogger) writing about a typhoon in the Philippines. In the less severe deception condition, the scenario stated that the photo was "*taken of the same typhoon, but when it was hitting a neighbouring country*". In the severe deception condition, the scenario stated that the photo was "*however, taken of a different typhoon hitting the same area a few years ago*".

Each scenario was presented in the same order in each HIT and was paired with an image. All photos were identical across all conditions, except in the last scenario concerning O.J. Simpson, where the severe deception scenario contained the fact that the image had been altered, and a photo-shopped version of the image was displayed.

Participants were provided with a question after every scenario-image pair, which stated "*This is the only image you have. Do you use this image?*" This was followed by '*yes*' or '*no*' radio buttons and a text box asking participants to type the reasons for their decisions.

Analysis

After all data had been gathered, qualitative data was coded using an axial coding method. The results of this coding were used to create categorical variables. The effect of this, as well as the effects of role and severity of deception on decisions to use images, were analysed using Pearson Chi-Square tests in SPSS.

Results

No significant effects were found for either the role or degree of deception on the decisions participants made about the hypothetical use of images. A marginally significant effect was found in the OJ Simpson scenario where participants were slightly more likely to click 'no' to using the image in the severe deception condition than in the slight deception condition, $t(1, 118) = 2.853$, $p = .094$.

Sixty-five% of participants in all conditions opted to use image 1, 59% for image 2 and 85% said that they would use image 3. By contrast, only 38% decided they would use image 4.

As a result of qualitative coding, three themes emerged in which participants were basing their decisions; upon *ethical* considerations, *compromise*, or talking in terms of *cost vs. gain*.

	Ethical	Cost vs. Gain	Compromise
1	34.1	35.22	30.68
2	42.11	18.42	39.47
3	48.78	12.2	39.02
4	43.89	13.26	42.85

Table 2 – Percentages of decision basis used by image

A chi-square analysis revealed that whether or not participants talked about cost vs. gain had a significant interaction with levels of trustworthiness judged in the OJ Simpson scenario, $c^2(1, 120) = 8.53$, $p = <.01$. When participants spoke in terms of cost vs. gain in their reasoning for their answers to this scenario, they were significantly more likely to decide not to use the deceptive image than if they did not talk about consequences. This was compared to no difference found when participants did not talk in terms of cost vs. gain.

Discussion

Across both deception conditions and roles, the majority of participants chose to deceive with the three images, with the result of the fourth image being reversed for reasons we will discuss below. This is perhaps suggestive of a propensity for human subjects to deceive with images.

The lack of a significant result in the difference between roles could have been due to a variety of factors. (Smith, Houwer & Nosek, 2013, Pornpitakpan, 2004 & Davis, 2008), suggests that there should have been an effect, and to say that our sample-size was simply too small, or that the role-induction was ineffective would be plausible as an explanation. However, people's perceptions of the credibility and trustworthiness of journalists and bloggers appear to be more complex than we first anticipated. For example, Mackay & Lowrey (2011) found conflicting results investigating the effects of different types of journalists on trustworthiness, and found a greater effect in the attachment of journalists to a particular institution. There was an emphasis on political views in their experiments, so the perceived political ties of each media institution used may have accounted for part of the effect. Some of our images had political connotations and some didn't. The fact that people may have been relying on different indicators to base the actions of their role may have confounded the results. In fact, Scholl & O'Hair (2005) found that a number of personal attributes and beliefs contribute to a person's propensity to deceive, and, judging by how ingrained they are, it may mean that any superficial attempt at applying a role would prove futile in swaying these beliefs and may explain why the reasons participants gave were more indicative of the answers than the roles.

Despite the logical conclusion that journalism would be perceived as more trustworthy given its heightened credibility, regulated professional standards, and strict accountability, the actual perceptions of the participants in this study may not

have reflected this. One study (Johnson & Kaye, 2004) shows that people who regularly use and interact with blogs tend to rate their credibility and truthfulness as considerably higher than that of traditional journalism. It is highly likely that the nature of the recruitment and participation of our participants, being exclusively web-based, may be skewed towards a primarily blog-using sample. This may have minimalised the difference in credibility perceptions of bloggers and journalists, and, coupled with many participants not appearing to fully inhabit the roles they were given, rendered the effects of these roles ineffective. With this in mind, any future studies wishing to use journalists and bloggers as means to induce disparities in perceived trustworthiness should be mindful of the preferences towards online or traditional reporting of their participants, and possibly, may need to adjust their predictions accordingly.

Correlations between participants' answers and their likelihood to speak in terms of cost and gain were almost non-existent, except in the OJ Simpson condition, where the likelihood to discuss cost versus gain significantly correlated with decisions not to endorse the images. The fact that this only occurs in the OJ Simpson condition relates to the fact that people are less likely to deceive when the cost of being found out is greater (Gneezy, 2005). This makes sense in the context of the present study as, in this condition, the cost was designed – and perceived – to be greatest. The famous nature of the subject matter meant that a deception would be greatly more recognisable than in the other conditions where one would have to be familiar with the difference in flora in Asia, or the demographics of tourists visiting Machu Picchu to discover a deception in the same way people would by looking at the face or time of the OJ Simpson mug shot.

Percentages of yes or no responses for Images 1, 2 and 3 revealed that participants were more likely to agree to using an image than not. This may have been due to the perceived non-seriousness of the deceptions in these conditions. However, even in the photoshopped level of the OJ Simpson condition, considering the recognisability of the deception as well as the fact that it had been manipulated for the purpose of misleading hypothetical readers, there were still 38% of participants endorsing the image. This further shows the tendency toward deception in endorsing slightly untruthful images that we saw in the earlier conditions is carried through – in albeit smaller numbers – to a dramatically misleading form of deception.

Broader implications of findings of Experiment 1 and 2

The dimensions of trustworthiness identified in Experiment 1 allow a comparison of user-centred themes of trustworthiness, which were induced from a qualitative study of archival documents (Donaldson & Conway, 2015). This study revealed the following dimensions:

- Accuracy: believed to be free of error
- Believability: the extent to which the information appears to be plausible
- Coverage: completeness of the information
- Currency: the degree to which the information is up-to-date
- Objectivity: balance of content
- Stability: the persistence of information, both its presence and contents
- Validity: the use of responsible and accepted practices such as the soundness of the methods used, the inclusion of verifiable data, and the appropriate citation of sources

The following emergent themes were additionally uncovered:

- Perceived authenticity: Is it fake?
- Inaccurate information: conceptualizing documents as being trustworthy despite containing inaccurate information
- Primary or first hand evidence: the extent to which the document is primary or first-hand
- Document legibility or readability
- Document's perceived proper form.

Although the preceding themes "coverage", "readability", "proper form" and "validity" relate to the information being in the form of a document, it is nevertheless possible to draw some comparisons with the four dimensions that were induced from Experiment 1 based on information in the form of images.

Donaldson & Conway's theme of "authenticity" relates to the "image features" dimension as the latter comprises the identification of areas of the image that look fake or suspicious. A contrast can also be drawn with the theme of "accuracy", but in the case of images it is not freedom from error that underpins the decision but issues such as whether the image is deemed an "accurate enough" portrayal of the subject of the image.

In both documents and images "believability" hinges on determining the plausibility of the content. Experiment 1 uncovered that in the case of images prior knowledge is an important component of that determination.

An important and unexpected finding from experiment 1 is that the subject of an image has a considerable effect on the judgement of trustworthiness of that image, particularly if the viewer distrusts the subject. In some ways this finding is the converse of an aspect of the "validity" uncovered by Donaldson & Conway where a document would be regarded as trustworthy even if its content is inaccurate, or incorrect (i.e., untrustworthy). In experiment 1, an image could be deemed *untrustworthy* simply because the subject (i.e., the content) of the image is deemed untrustworthy. If this finding is reliable, it has implications for using people to judge the credibility of images, as it appears that for some images the decision about the trustworthiness of the image is being confounded with a decision regarding the trustworthiness of the *subject* of the image.

Findings from both experiments 1 and 2 represent a challenge to the cultural heritage sector. They suggest that it cannot be assumed that digital images from or of their collections will necessarily or automatically be trusted by those viewing these artefacts. Library, archival and museum advocacy and outreach programmes concentrating on increasing the visibility of collections by curating online exhibitions need to develop awareness of the nuanced and complex factors influencing user decision making about trust issues. This is a particularly significant issue for the archival professional community given the critical nature of ensuring the trustworthiness of records (see, for example, Duranti and Rogers, 2012). The findings from the two experiments suggest that it is overly simplistic to assume that by establishing a link with a cultural heritage institution trust will automatically ensue.

The relative trust accorded to journalists and bloggers also suggest that trust once accorded to established or traditional professions should not be assumed to be the case in today's digital environment. This calls into question the extent to which it can be assumed that cultural heritage professionals (librarians, archivists, museum curators) will be automatically assumed to confer trustworthy status on digital artefacts. Overall, the findings suggest the need for much more research in this area, comparing for instance the impact of making collections available on different platforms.

Conclusion

This article set out to explore the cognitive decision space for deciding the trustworthiness of images. The qualitative analysis from Experiment 1revealed the following four themes, which we put forward as corresponding to underlying dimensions of this decision space. This parallels research in document relevance which revealed dimensions such as "topical relatedness", "novelty", etc.:

- Features of the image itself (e.g. '*The person in the foreground does not seem to blend with the image in the background'*)
- Content/Subject of the image (e.g. '*All wild animals are untrustworthy. Also look at his teeth. He acts on instinct alone.'*)
- Image source (e.g. *'it is from a museum, so it seems to be trustworthy'*)
- Prior knowledge (e.g.*' Could be faked, but the auroras are spectacular and always look fake.'*)

Studies into the credibility of media have generally been divided into two areas, source credibility (concerning the individual bearing information) and medium credibility (concerning the wider entity through which information is broadcast). In the present study, the focus of experimental manipulations was on the latter form. Two sides of this form of credibility were addressed; media as institutions (museums versus Facebook as in Experiment 1), and media in terms of types of reporting medium (journalists versus bloggers). The presumption here is that the new forms of media are less bound by the rules placed on the traditional, being born in an era where artistic licence is given greater precedence as entertainment.
One of the stand out results of these experiments are that, even though there seems to be a different degree of trust between sources of distinct levels of credibility, the same distinction does not occur when asked to inhabit sources and transact decisions of trust. This suggests that the standards people place on others are disparate to those that they place on themselves.

Acknowledgements
The authors are grateful to the InterPARES Trust (https://interparestrust.org) which funded this research.

References

Alter, A. L., & Oppenheimer, D. M. (2009). Uniting the Tribes of Fluency to Form a Metacognitive Nation. *Personality and Social Psychology Review, 13*(3), 219-235. doi:10.1177/1088868309341564

Barry, C. (1994). User-defined relevance criteria: an exploratory study. *J. Am. Soc. Inform. Sci.* 45, 145–159.

Borlund, P. (2003). The concept of relevance in IR. *J. Am. Soc. Inform. Sci. Tech.* 54, 913–925. doi: 10.1002/asi.10286

Bosman, J. (2004). First with the Scoop, If Not the Truth. *New York Times, April 18, 2004, section 9*, p. 10.

Caldwell, S., Gedeon, T., Jones, R. and Copeland, L. (2015). "Imperfect Understandings: A Grounded Theory and Eye Gaze Investigation of Human Perceptions of Manipulated and Unmanipulated Digital Images" in *Proceedings of the World Congress on Electrical Engineering and Computer Systems and Science (ECSS 2015)*

Carlson, M. (2009). The Reality of a Fake Image; News norms, photojournalistic craft, and Brian Walski's fabricated photograph. *Journalism Practice*, *3*(2), 125–139.

Cattell, R. B., & Cattell, H. E. (1995). Personality structure and the new fifth edition of the 16PF. *Educational and Psychological Measurement, 55*(6), 926-937.

Coleman, S. E. (2007). *Digital photo manipulation: A descriptive analysis of codes of ethics and ethical decisions of photo editors* (Ph.D.). The University of Southern Mississippi, Ann Arbor. Retrieved from ProQuest Dissertations & Theses Global. (304825651)

Cook, K. S., & Gerbasi, A. (2009). Chapter 10: Trust. In P. Hedström & P. S. Bearman (Eds.), *Oxford Handbook of Analytical Sociology* (pp. 218-241): Oxford University Press.

Cooper, C. (2015). *Individual Differences and Personality* (3 ed.). London: Taylor and Francis.

Davis, R. (2008). A symbiotic relationship between journalists and bloggers. *Joan Shorenstein Center on the Press, Politics and Public Policy, John F. Kennedy School of Government, Harvard University*.

Donaldson, D.R & Conway, P. (2015). User conceptions of trustworthiness for digital archival documents. *Journal of the Association for Information Science and Technology*, 66(12), 2427-2444.

Doney, P. M., & Cannon, J. P. (1997). An examination of the nature of trust in buyer-seller relationships. *The Journal of Marketing*, 35-51.

Duranti, L., & Rogers, C. (2012). Trust in digital records: An increasingly cloudy legal area. *Computer Law & Security Review*, *28*(5), 522–531.

Eysenck, H. J. (1992). Four ways five factors are not basic. *Personality and individual differences*, *13*(6), 667-673.

Gneezy, U. (2005). Deception: The role of consequences. *American Economic Review*, 384-394.

Greenberg, G. (2013). Beyond resemblance. *Philosophical Review* 122(2): 215-287

Halberstadt, J., & Rhodes, G. (2003). It's not just average faces that are attractive: Computer-manipulated averageness makes birds, fish, and automobiles attractive. *Psychonomic Bulletin & Review*, *10*(1), 149-156.

InterPARES. (2008). InterPARES 3 Project: International Research on Permanent Authentic Records in Electronic Systems. Retrieved July 22, 2014, from http://www.interpares.org/ip3/ip3_terminology_db.cfm?letter=t&term=55

Jacoby, L. L., Kelley, C. M., & Dywan, J. (1989). Memory attributions. In H. L. Roediger & F. I. M. Craik (Eds.), *Varieties of memory and consciousness: Essays in honour of Endel Tulving* (pp. 391-422). Hillsdale, NJ: Erlbaum.

Johnson, T. J., & Kaye, B. K. (2004). Wag the blog: How reliance on traditional media and the Internet influence credibility perceptions of weblogs among blog users. *Journalism & Mass Communication Quarterly*, *81*(3), 622-642.

Kelton, K., Fleischmann, K.R., & Wallace, W.A. (2008). Trust in digital information. *Journal of the American Society for Information Science and Technology*, 59(3), 363–374.

Kim, J., Kazai, G., and Zitouni, I. (2013). "Relevance dimensions in preference-based IR evaluation", in *Proceedings of the 36th Annual ACM Conference of Research and Development in Information Retrieval (SIGIR' 13)* (New York, NY: ACM Press), 913–916.

Los Angeles Times. (2005). Los Angeles Times Ethics Guidelines. Retrieved from http://media.trb.com/media/acrobat/2005-07/18479691.pdf

Lum, J. (2010, July 5). Controversy Crops Up Over Economist Cover Photo. PetaPixel. Retrieved from http://petapixel.com/2010/07/05/controversy-crops-up-over-economist-cover-photo/

Lum, J. (2010, July 19). Getty Photographer Terminated Over Altered Golf Photo. PetaPixel. Retrieved from http://petapixel.com/2010/07/19/getty-photographer-terminated-over-altered-golf-photo/

Mackay, J. B., & Lowrey, W. (2011). The credibility divide: reader trust of online newspapers and blogs. *Journal of Media Sociology*, *3*(1-4), 39-57.

MacNeil, H. (2001). Trusting records in a postmodern world. *Archivaria*, 36–47.

Martindale, C., & Moore, K. (1988). Priming, prototypicality, and preference. *Journal of Experimental Psychology: Human Perception and Performance*,*14*(4), 661.

Mizzaro, S. (1997). Relevance: the whole history. *JASIS.* 48(9), 810–832.

National Press Photography Association. (2012). NPPA Code of Ethics. Retrieved May 19, 2014, from https://nppa.org/code_of_ethics

Nissenbaum, H. (2001). Securing Trust Online: Wisdom or Oxymoron. *Boston University Law Review, 81*(3), 635-664.

Popper, K. R. (1972). *Objective knowledge: An evolutionary approach*. Oxford: Oxford University Press.

Pornpitakpan, C. (2004). The persuasiveness of source credibility: A critical review of five decades' evidence. *Journal of Applied Social Psychology*, 34(2), 243-281.

Reber, R., & Schwarz, N. (1999). Effects of perceptual fluency on judgments of truth. *Consciousness and Cognition, 8*(3), 338-342.

Repp, B. H. (1997). The aesthetic quality of a quantitatively average music performance: Two preliminary experiments. *Music Perception*, 419-444.

Roberts, P., & Webber, J. (1999). Visual truth in the digital age: Towards a protocol for image ethics. *Australian Computer Journal*, *31*(3), 78–82.

Schamber, L., Eisenberg, M., and Nilan, M. (1990). A re-examination of relevance: toward a dynamic, situational definition. *Information Processing & Management.* 26(6), 755–775.

Scholl, J. C., & O'Hair, D. (2005). Uncovering beliefs about deceptive communication. *Communication Quarterly*, *53*(3), 377-399.

Shah, A. K., & Oppenheimer, D. M. (2007). Easy does it: The role of fluency in cue weighting. *Judgment and Decision Making, 2*(6), 371-379.

Sofer, C., Dotsch, R., Wigboldus, D. H., & Todorov, A. (2015). What is typical is good: The influence of face typicality on perceived trustworthiness. *Psychological Science*, 26(1), 39-47. doi:10.1177/0956797614554955

Swanson, D. R. (1986). Subjective versus Objective Relevance in Bibliographic Retrieval Systems. *The Library Quarterly: Information, Community, Policy, 56*(4), 389-398. Retrieved from http://www.jstor.org/stable/4308045

Taddeo, M. (2009). Defining Trust and E-Trust: From Old Theories to New Problems. *International Journal of Technology and Human Interaction (IJTHI), 2*(5), 23-35. doi:10.4018/jthi.2009040102

Taddeo, M., & Floridi, L. (2011). The case for e-trust. *Ethics and Information Technology, 13*(1), 1-3. doi:10.1007/s10676-010-9263-1

Todorov, A. (2008). Evaluating faces on trustworthiness. *Annals of the New York Academy of Sciences*, *1124*(1), 208-224.

Todorov, A., & Duchaine, B. (2008). Reading trustworthiness in faces without recognizing faces. Cognitive Neuropsychology, 25(3), 1-16.

doi:10.1080/02643290802044996

Willis, J., & Todorov, A. (2006). First impressions making up your mind after a 100-ms exposure to a face. *Psychological science*, *17*(7), 592-598.

Winkielman, P., Halberstadt, J., Fazendeiro, T., & Catty, S. (2006). Prototypes Are Attractive Because They Are Easy on the Mind. *Psychological Science, 17*(9), 799-806.

Winkielman, P., Schwarz, N., Reber, R., & Fazendeiro, T. A. (2003). Affective and Cognitive Consequences of Visual Fluency: When Seeing is Easy on the Mind. In R. Baatra & L. Scott (Eds.), *Persuasive imagery: A consumer response perspective* (pp. 75-89). Mahwah, NJ: Lawrence Erlbaum.